Journal of Surgical Oncology 1998;67:143–150

GUEST EDITORIAL

The Breast Cancer Screening Controversy: Lessons to Be Learned

DANIEL B. KOPANS, MD*
Harvard Medical School, Massachusetts General Hospital, Boston, Massachusetts

INTRODUCTION

In April 1997, the National Cancer Institute (NCI) reversed the position it had articulated in 1993 and, once again, voiced support for screening women ages 40–49 for breast cancer. This ended a 4-year debate that had been caused by inappropriate scientific analyses and supported by the misleading presentation of data.

The age of 50 has no biological significance. None of the parameters of screening changes abruptly at the age of 50. The recall rates for an abnormal mammogram are approximately the same, regardless of age, as are recommendations for biopsy initiated by mammography. The cancer detection rate increases steadily with increasing age paralleling the prior probability of breast cancer in the population, with no abrupt change at age 50 [1]. There is statistically significant proof, from randomized, controlled trials, that periodic screening, using mammography, can reduce the death rate from breast cancer for women beginning by the age of 40.

The screening controversy began as a consequence of the inappropriate use of subgroup analysis that lacked statistical power in the formulation of medical recommendations and was prolonged by a marked imbalance of publications in important medical journals that prevented the issues from being presented to physicians and the public in a balanced and impartial fashion. Physicians should be aware of the importance of appropriate data analysis and should insist on balanced presentations of controversial topics in their journals.

HISTORICAL BACKGROUND

One of the most contentious issues in health care over the past several years has been the question of at what age women should be advised to begin screening for breast cancer using mammography. The debate had actually been simmering for more than 20 years, but the seeds of the recent argument were sown in a review by Eddy et al. [2] in the *Journal of the American Medical Association* in 1988. In that article the authors stated that there was a benefit from screening women in their forties, but that these women should expect to pay for it themselves. Many of the arguments developed in that article have been repeated in the debate of the past 4 years. Among them are the ''harms'' of screening that include the anxiety of having a mammogram, the anxiety caused by a false-positive report, and the trauma and anxiety from a biopsy with benign results. Motivations are difficult to assess, but King [3] suggested that it was a debate over public health policy, stating ''Screening is a public health activity concerned not with individuals, but with populations . . . ,'' while many proponents see screening as a way to decrease deaths from breast cancer among individuals, with the only public health issues revolving around access to screening and reimbursement [4].

During the 1980s, many differing guidelines for breast cancer screening had been promulgated by the major medical organizations. In the spring of 1989, in an effort to reduce the confusion, nine major organizations including the NCI, the American Medical Association (AMA), and the American Cancer Society (ACS), agreed on what became known as the ''consensus guidelines'' that recommended a baseline mammogram for ages 35–39 (the baseline was later dropped due to lack of scientific support), mammography every 1–2 years for women in their forties, and then annual mammography for women age 50 and older.

Just 3 years later, the preliminary results from the National Breast Screening Study of Canada (NBSS) were published suggesting that there was no benefit from screening women ages 40–49 [5]. Despite major concerns about the design and performance of the Canadian study [6], the leadership at the NCI convened a work-

*Correspondence to: Daniel B. Kopans, MD, Department of Radiology, Massachusetts General Hospital, 55 Fruit St., Boston, MA 02114. Fax No.: (617) 726-1074.

Accepted 28 November 1997

shop, in February 1993, to re-evaluate the available data. Arguments boiled to the surface after the ''International Workshop on Screening of Breast Cancer.'' The NCI was criticized for orchestrating a *fait accomplis* by inviting only three experts to argue in support of screening women ages 40–49 while inviting 17 others to argue against screening these women [7]. In addition, four out of the five authors, chosen by the NCI to write a report summarizing the Workshop, had already written in opposition to screening.

The Workshop report was published in October 1993 and concluded that there were no data to support screening women ages 40–49 [8]. Only data from the eight randomized controlled trials (RCT) of screening were evaluated. Other data were excluded. What the report failed to point out was that the trials, when analyzed as they were designed, demonstrated a statistically significant benefit for screening women beginning by the age of 40 [9]. The misinterpretation was a result of the inappropriate use of subgroup analysis in the assessment of the data. Despite the fact that the trials were not designed to permit subgroup analysis, the data were, retrospectively, stratified by age separating the results for women ages 40–49 from the overall data analysis. Even though 5 of the 8 trials in 1993 were showing a benefit for the younger women, when women in their forties were evaluated separately, the benefit did not reach statistical significance, and it was dismissed.

What the report failed to explain was that, since the trials were never intended to be analyzed by age subgroups, they lacked the statistical power, in the early years of follow-up, to permit legitimate subgroup analysis for women under the age of 50. Despite having been apprised of this [10], the authors of the Workshop report overlooked the fact that, even when the trials were all grouped together in a meta-analysis, it was mathematically impossible for an anticipated benefit of 25% to be statistically significant in the early years of follow-up [11] (see Appendix). Even the Canadian study, which had been promoted as having been ideally designed to analyze women ages 40–49 actually involved only one tenth the number of women needed to show an early benefit [12,13]. In addition, there were major problems with the poor quality of the mammography in the NBSS. Questions persist over the unblinded randomization in the NBSS, and the investigators have yet to explain satisfactorily why, even with 10 years of follow-up evaluation, there are still many more women with advanced breast cancers and cancer deaths in the screened group relative to the controls, when none of the other trials has had this result [14].

Recognizing the major flaws in the NCI analysis, the National Cancer Advisory Board (NCAB), in 1993, voted 14–1 to advise that the NCI not change the guidelines. Despite this vote, and for the first time in the history of the NCI, the Director ignored the advice of the NCAB and, in December of 1993, withdrew NCI support for screening women ages 40–49.

WHY AGE 50?

Since the development of breast cancer is clearly hormone related, it is not unreasonable to wonder whether menopause has an influence on cancer detection and prognosis. The first randomized, controlled trial of screening, undertaken within the Health Insurance Plan of New York (HIP), decided to see if menopause had any influence on the trial results and the age of 50 was chosen as a surrogate for menopause. The data were analyzed retrospectively by grouping the participants by ages 40–49 and ages 50–64. A statistically significant benefit was found for the older women, when analyzed separately, but the benefit for younger women did not quite achieve significance, even though it was almost as great as for women ages 50–64 [15]. Consequently, some investigators concluded that there was no benefit for women in their forties.

Subsequently, the other screening trials were also, retrospectively evaluated by grouping the data for women 49 and younger separately from those age 50 and older. It had been pointed out to the NCI, during the 1993 Workshop, that recommendations should not be based on subgroup analysis of the trials since they lacked the statistical power to permit this. Even if all the world's trials were grouped together, they had actually included only one third the number of the women needed to ''prove'' a statistically significant benefit in the short-term follow-up results [16]. Although this was dismissed by the NCI in 1993, it was quietly acknowledged subsequently, in 1994. Rather than clarify the mistake, the NCI proposed a new, larger trial that would have sufficient power [17]. Subgroup analysis of data that lack statistical power is used to determine the next research question, but it can be so misleading that it should not be used to guide medical recommendations (see Appendix).

There has actually never been good biological support for dividing the trial data at ages 49 and 50, but the issue became one of public health planners concerned about the cost of screening while physicians were trying to determine whether to recommend screening to their patients and women were trying to decide whether or not to be screened.

The controversy that followed the NCI decision in 1993 triggered a Congressional review that was highly critical of the NCI decision. It termed the guidelines change ''Misused Science'' [18]. Among the criticisms lodged at the NCI by the Congressional review was the fact that it had been incomplete and biased, that it had not pursued the proper format for making such an important decision, and that, in the future, such reviews should be done by a consensus conference.

JOURNAL PUBLICATION BIAS

The controversy was sustained by numerous journal publications in general medical journals as well as specialty journals in support as well as opposition to screening women in their forties. However, in the major journals that are monitored by the media, there was a great preponderance of publications that supported the NCI opposition to screening. From 1993 through 1996, numerous submissions of material that provided data and analysis in support of screening women ages 40–49 were made to several of the leading medical journals, the *New England Journal of Medicine,* the *Journal of the American Medical Association,* the *Journal of the National Cancer Institute,* and the *Annals of Internal Medicine.* These articles in support of screening women ages 40–49 were routinely rejected while these journals published articles that either directly, or, indirectly argued in opposition to recommending screening to women in their forties repeatedly providing their readers and the public with arguments against screening women in their forties. Between 1993 and the end of 1996, there were 17 articles published in these journals [19–35] that favored opposition to screening women ages 40–49, while only three articles (letters to the editor are excluded, since they carry no scientific weight) were published supporting screening women ages 40–49 [36–38] despite numerous submissions including a review by eight of the leading experts on breast cancer screening in the United States. Of the three supporting articles published, two were limited, by the journals, to 2,500 words [36,37], while the opposition articles had what amounted to unlimited space [24,27]. Since these journals are among the publications most often cited by the media, the public and physicians were saturated with articles that opposed screening for younger women.

INAPPROPRIATE DATA GROUPING IS MISLEADING

During this time, many of the articles that opposed screening women in their forties, repeatedly presented the data by combining the results for women ages 40–49 as if they were a uniform group and comparing them to *all* women ages 50 and older, as if they were a uniform group. This fostered the idea that there was an abrupt change in many screening parameters that occurred at the age of 50. An example of this was a review, cited by the NCI in its argument for the 1993 guidelines change [39], that suggested that there was an abrupt change in the detection rate of breast cancers at the age of 50 [16]. Despite having been alerted to the fact that the data actually did not show any abrupt change at the age of 50, but clearly showed a steady increase in the cancer detection rates for women between the ages of 30 and 70+ (D.B. Kopans to K. Kerlikowske, personal communication, 1993), the data were grouped together. A dichotomous comparison was made comparing women ages 30–49 as if they were a uniform group, with all women ages 50 to 70+ as if they were a uniform group. This created the misleading appearance of an abrupt change in cancer detection rates at the age of 50 [23], when there was none. A careful review of many of the arguments put forth during that period, and continuing today, by those opposed to recommending screening for women ages 40–49, shows the repeated use of similar dichotomous analyses that compared women ages 40–49 to all women over the age of 50. When this type of dichotomous comparison is made, any factors that change gradually with increasing age appear to change abruptly at the age of 50. In fact, none of the parameters of screening changes abruptly at any age [1].

THE 1997 CONSENSUS DEVELOPMENT CONFERENCE

In 1995, a new director was appointed to head the NCI. Informed of the problems with the 1993 decision (D.B. Kopans, M.D. to R. Klausner, M.D., personal communication), he was apprised of the most recent data from the randomized, controlled trials of screening in the Spring of 1996. These demonstrated that, with longer follow-up and larger numbers, there was a ''statistically significant'' benefit from screening women ages 40–49 even when they were segregated by age [40]. Seeking an unbiased review, the new Director requested a Consensus Development Conference (CDC) to review the subject and this was organized by the Office of Medical Applications of Research (OMAR) within the NIH.

A CDC is intended to be an objective review of the available data by disinterested experts who have not been involved in the debate in question. There were a considerable number of concerns raised during the development of the agenda for the CDC. These included significant conflicts of interest and an imbalance in presentations as well as the secret selection of a, potentially, biased panel (D. Kopans, M.D., to R. Klausner, M.D., J.M. Elliot, and J.H. Ferguson, M.D., personal communications). By its design, a CDC is similar to a jury trial, where proponents and opponents provide the arguments for and against screening. A panel is chosen to act as the jury, to weigh the evidence and to reach a consensus. The panel is supposed to be objective, and to have no previous involvement in the controversy under discussion and should have no conflicts of interest. This proved not to be the case.

Although the panel was termed ''expert,'' many of the panel members were chosen specifically because they lacked expertise with regard to the topic (to avoid any preconceived positions), while others had significant, potential, conflict of interest including direct associations with major opponents of screening as well as significant

funding from NCI, whose policy was under review. These conflicts of interest were not disclosed.

The panel convened in January 1997 and heard a day and a half of testimony. Despite having provided reassurances that discussion would be thorough and complete, many of the discussions were cut short. The charge to the panel had been to review the latest data from the randomized, controlled trials (RCTs) and to take into account data from other large trials, even if they were not RCTs. At the culmination of the CDC meeting, the panel generated a summary that was released to the meeting and to the media. The panel statement concluded that the available data did not support a recommendation for screening women ages 40–49 [41].

The audience, especially the Swedish investigators whose latest data demonstrated a clear benefit from screening women ages 40–49, were stunned by the panel summary. The trialists from the screening programs had spent a great deal of effort collecting and analyzing the latest data from their trials and had presented them to the panel. The panel was specifically charged with evaluating the most recent information yet, the panel, in their statement, *never mentioned any of the new data.* In addition, the panel report, issued January 23, 1997, made statements that were misleading and factually incorrect.

The panel never pointed out that the trial data were not intended for subgroup analysis and that subgroup analysis could be misleading because the trials lacked statistical power during the early years of follow-up evaluation (see above).

The panel never provided women and their physicians the new screening trial results including the fact that the Gothenberg trial reported a 44% mortality reduction that was statistically significant by itself [24]. The Malmo trial, by itself, had a statistically significant mortality reduction of 35% [43].

The panel was inconsistent when they overlooked the fact that there are only two trials which, by themselves, have been significant for women ages 50 and older, and yet they reinforced support for screening beginning by the age of 50.

The five Swedish trials (the most similar in design) had a 29% reduction from screening women in their forties that was statistically significant [44]. When the results from the HIP and Edinburgh trials are added to include all of the trials by ''invitation'' to be screened, there is a 26%, statistically significant benefit [45]. Even when the negative results from the NBSS are included, despite the fact that it was totally different from the other trials and has major, unexplained problems, there is a benefit of 18% that is statistically significant. A statistically significant benefit was the requirement set by the NCI in 1993. The fact that it had been met was all the more remarkable given the weak statistical power of the subgroup analysis for women ages 40–49.

The panel statement included a series of factually incorrect statements. It suggested that if there were any benefit for women ages 40–49, it was no greater than 30%. Clearly, this was not supported by the facts given that Gothenberg had a statistically significant, 44% mortality reduction and Malmo had a 35% reduction.

The panel statement made a point of the fact that there was no significant benefit before 7 years. The panel never explained why they required a benefit by 7 years. This was never a stated goal of any of the trials. Furthermore, the statement suggested that, if there was any benefit, it took at least 10 years to appear, when the facts contradicted this. A benefit in RCT can only occur for the screened women when control women die (see Appendix). The panel ignored the fact that the screening intervals were not optimized to interrupt faster growing cancers. Most had two years between screens. This meant that only moderate growth cancers could be interrupted. Younger women, stage for stage, live longer with their cancers than older women [46], and so it is not surprising that it would take a little longer for the benefit to appear. Nevertheless, in three of the trials, HIP, Malmo, and Gothenberg, the benefit began to appear 5–7 years after the onset of screening, contradicting the panel statement.

The panel statement suggested that if there was a benefit it might be due to cancers detected by clinical breast examination in the trials. This was clearly not supported by the facts. The Swedish trials, which, in addition to Malmo and Gothenberg, with their separate benefit, collectively, had a 29% benefit that is statistically significant, *did not provide a clinical breast examination.*

The panel statement suggested that mammography misses 25% of cancers in women ages 40–49, while it only misses 10% of cancers in women age 50 and older. The Panel failed to point out that they had been provided data showing that modern mammography, performed every year, detects 85% of cancers among women in their forties [47,48].

The panel statement suggested that the benefit from screening women ages 40–49 could well be due to the fact that they reached age 50 during the trials, and screening began to work. This has no biological basis, and is contradicted by the facts. Just as it is required that women who die of breast cancer having been offered screening, but refused must still be counted as having been screened (noncompliance), it is well established that RCT must be analyzed by the age at allocation and not the age at diagnosis to avoid significant biases (see below). Nevertheless, the data show that, in the three trials that have analyzed their data by the age at diagnosis, the benefit was, predominantly, from cancers detected while the women were in their forties [49–51].

NCI DISREGARDS THE CONCLUSIONS OF THE CDC PANEL

The director of the NCI immediately registered his disagreement with the panel's conclusions having, himself, heard the latest data and discussions at the CDC. One of the panel members resigned in protest over the statement, and two other members, subsequently, wrote a minority report that supported screening women in their forties [52]. The NCAB was asked to review the data. On the basis of this review, 2 months later, the NCI, once again gave its support for screening women in their forties. The ACS also reviewed the data and reaffirmed the ACS support for screening women ages 40–49. Unlike the NCI, which returned to the option of every 1–2 years for women ages 40–49, the ACS urged women to be screened every year, since there was increasing evidence that cancers among younger women grow faster, and that the time between screens should be shorter to interrupt their growth before they become successfully metastatic.

LESSONS TO BE LEARNED FROM THE CONTROVERSY

The issues involved in this controversy are not particularly unique and might apply to any number of controversial topics in health care.

Statistical Power Is Not Trivial

The main issue has been defining the appropriate analysis of data from randomized, controlled trials. There are established guidelines for analyzing these data that have been developed over the years to avoid misinterpretation of the data and biases that can produce exactly the problems that have occurred.

The statistical power of the trials and the data analyzed, is critical. Lachin [53] warned that:

> If the statistical test fails to reach significance, the power of the test becomes a critical factor in reaching an inference. It is not widely appreciated that the failure to achieve statistical significance may often be related more to the low power of the trial than to an actual lack of difference between the competing therapies. Clinical trials with inadequate sample size are thus doomed to failure before they begin and serve only to confuse the issue of determining the most effective therapy for a given condition.

This was reinforced by Moher et al. [54]:

> If a trial with negative results has insufficient power, a clinically important but statistically nonsignificant effect is usually ignored, or worse, is taken to mean that the treatment under study made no difference.

These are not merely academic issues but bear directly on the critical analysis of data and the legitimate conclusions that can be drawn from the data. If statistical power is not important, then trials would only require a few participants. Clearly this is not the case.

Guidelines for Analyzing Randomized, Controlled Trial Data Can Help Avoid Misinterpretations

Noncompliance and contamination. There are other ''rules'' that must be followed to avoid biases. In seven out of the eight trials, women were first randomly divided and then were invited to be screened. Women who were allocated to the screening arm, but refused the invitation (noncompliance), must still be counted as having been screened. Should they die from breast cancer their deaths are counted as deaths among screened women. Similarly, a woman assigned to the control group who has her life saved as a result of a mammogram that she obtains outside of the trial (contamination) is still counted as an unscreened control. These rules are needed to avoid biasing the results. Nevertheless, they also mean that the trial results will underestimate the benefit from screening.

Age at allocation versus age at diagnosis. Another requirement of RCT analysis is that women must be evaluated by their age at allocation, not their age at diagnosis. Proroc et al. [55] pointed out that using the age at diagnosis can bias the conclusions because it represents a pseudo-variable that is influenced by the intervention. If the age at diagnosis is used the results will be biased against screening younger women (see Appendix).

Data Grouping Can Be Misleading

Data are grouped all the time, but the reader should be informed as to how the grouping might influence the conclusions. The dichotomous analysis of the trial data gave the faulty impression that parameters of screening change abruptly at the age of 50 when they do not.

Publication Bias

Among the disheartening revelations in this controversy is the recognition of major biases in medical journal publications. Physicians are required to disclose any conflicts of interest, and major investigations have been launched to uncover scientific misconduct on the part of investigators, but medical journals have no ethical oversight. As demonstrated above, one need only review the publications of the major medical journals that are monitored by the media to realize that there has been a major imbalance in publications by several important journals in this controversy. Most recently, the *New England Journal of Medicine* (*NEJM*) twice turned down a fully documented, factually correct summary of the trial data

(some of which are presented above) that had been ignored by the CDC panel. This had been submitted in response to two *NEJM* articles, one of which suggested that the NCI decision in 1997 to once again support screening, was a political decision [56], while the second suggested that the decision whether or not to be screened was no more than a flip of a coin [57]. The facts were turned down without scientific comment.

CONCLUSION

There has been evidence for years that screening can benefit women beginning in their forties. There is now statistically significant ''proof.'' Several studies now suggest that cancers grow faster among younger women (where younger ends and older begins is unclear—it is not at age 50). As a consequence, the data suggest that younger women should be screened on an annual basis [58–60]. Going to a longer time between screens among older women may reduce the costs, but some faster growing cancers will likely not be intercepted in time. The balance between cost and benefit needs to be studied, but for now, the only reason to go to a longer time between screens is economic.

REFERENCES

1. Kopans DB, Moore RH, McCarthy KA, et al.: The positive predictive value of mammographically initiated breast biopsy: There is no abrupt change at age 50 years. Radiology 1996;200:357–360.
2. Eddy DM, Hasselblad V, McGivney W, Hendee W: The value of mammography screening in women under 50 years. JAMA 1988; 259:1512–1519.
3. King J: Mammography screening for breast cancer. Letter to the editor. Cancer 1994;73:2003–2004.
4. Kopans DB: Reply. Cancer 1994;73:2004–2006.
5. Miller AB, Baines CJ, To T, Wall C: Canadian National Breast Screening Study: 1. Breast cancer detection and death rates among women aged 40–49. Can Med Assoc J 1992;147:1459–1476.
6. Kopans DB, Feig SA: The Canadian National Breast Screening Study: A critical review. AJR 1993;161:755–760.
7. Kopans DB: Mammography screening and the controversy concerning women aged 40–49. Radiol Clin North Am 1995;33: 1273–1290.
8. Fletcher SW, Black W, Harris R, et al.: Report of the International Workshop on Screening for Breast Cancer. J Natl Cancer Inst 1993;85:1644–1656.
9. Shapiro S: Screening: Assessment of current studies. Cancer 1994;74:231–238.
10. Kopans DB: The National Breast Screening Study of Canada. A critical review of the results for women ages 40–49. Presented to the International Workshop on Screening of Breast Cancer, Bethesda, MD, Feb. 24–25, 1993.
11. Kopans DB: The National Breast Screening Study of Canada. A critical review of the results for women ages 40–49. Presented to the International Workshop on Screening of Breast Cancer, Bethesda, MD, Feb. 24–25, 1993.
12. Miller AB, Howe GR, Wall C: The National Study of Breast Cancer Screening. Clin Invest Med 1981;4:227–258.
13. Sickles EA, Kopans DB: Deficiencies in the analysis of breast cancer screening data. J Natl Cancer Inst 1993;85:1621–1624.
14. Tarone RE: The excess of patients with advanced breast cancers in young women screened with mammography in the Canadian National Breast Screening study. Cancer 1995;75:997–1003.
15. Shapiro S, Venet W, Strax P, Venet L: ''Periodic Screening for Breast Cancer: The Health Insurance Plan Project and its Sequelae, 1963–1986.'' Baltimore: Johns Hopkins University Press, 1988.
16. Kopans DB, Halpern E, Hulka CA: Statistical power in breast cancer screening trials and mortality reduction among women 40–49 with particular emphasis on the National Breast Screening Study of Canada. Cancer 1994;74:1196–1203.
17. Eyre H, Sondik E, Smith RA, Kessler L: Joint Meeting on the Feasibility of a Study of Screening Premenopausal Women (40–49 years) for Breast Cancer: April 20–21, 1994. Cancer 1995;75: 1391–1403.
18. House Committee on Government Operations: Misused Science: The National Cancer Institutes Elimination of Mammography Guidelines for Women in Their Forties. Union Calendar No. 480. House Report 103-863. October 20, 1994.
19. Kerlikowske K, Grady D, Barclay J, et al.: Positive predictive value of screening mammography by age and family history of breast cancer. JAMA 1993;270:2444–2450.
20. Kerlikowske K, Grady D, Rubin SM, et al.: Efficacy of screening mammography: A meta-analysis. JAMA 1995;273:149–153.
21. Kattiove H, Alessandro L, Keeler E, Brook RH: Benefits and costs of screening for early breast cancer: Development of a basic benefit package. JAMA 1995;273:142–148.
22. Lee Davis D, Love SM: Mammographic screening. JAMA 1994;271;152–153.
23. Fletcher SW, Fletcher RH: The breast is close to the heart. Ann Intern Med 1992;117:969–971.
24. Fletcher SW, Black W, Harris R, et al.: Report of the International Workshop on Screening for Breast Cancer. J Natl Cancer Inst 1993;85:1644–1656.
25. Baines CJ: The Canadian National Breast Screening Study: A perspective on criticisms. Ann Intern Med 1994;120:326–334.
26. Sox H: Screening mammography in women younger than 50 years of age. Ann Intern Med 1995;122:550–552.
27. Harris R, Leininger L: Clinical strategies for breast cancer screening: Weighing and using the evidence. Ann Intern Med 1995;122: 539–547.
28. Chalmers TC: Screening for breast cancer: What should national health policy be? J Natl Cancer Inst 1993;85:1619–1621.
29. Peer PG, Holland R, Jan HCL, et al.: Age-specific effectiveness of Nijmegen Population-Based Breast Cancer-Screening Program: Assessment of early indicators of screening effectiveness. J Natl Cancer Inst 1994;86:436–441.
30. Harris R: Breast cancer among women in their forties: Toward a reasonable research agenda. J Natl Cancer Inst 1994;86:410–412.
31. Kaluzny AD, Rimer B, Harris R: The National Cancer Institute and Guideline Development: Lessons from the breast cancer screening controversy. J Natl Cancer Inst 1994;86:901–903.
32. Swanson GM: May we agree to disagree, or how do we develop guidelines for breast cancer screening in women? J Natl Cancer Inst 1994;86:901–903.
33. Black WC, Nease RF, Tosteson ANA: Perceptions of breast cancer risk and screening effectiveness in women younger than 50 years of age. J Natl Cancer Inst 1995;87:720–731.
34. Rimer B: Putting the ''informed'' in informed consent about mammography. J Natl Cancer Inst 1995;87:703–704.
35. de Konig HJ, Boer R, Warmerdam PG, et al.: Quantitative interpretation of age-specific mortality reductions from the Swedish breast cancer-screening trials. J Natl Cancer Inst 1995;87:1217–1223.
36. Sickles EA, Kopans DB: Deficiencies in the analysis of breast cancer screening data. J Natl Cancer Inst 1993;85:1621–1624.
37. Sickles EA, Kopans DB: Mammographic screening for women aged 40–49 years: The primary care practitioners dilemma. Ann Intern Med 1995;122:534–538.
38. Lindfors KK, Rosenquist CJ: The cost-effectiveness of mammographic screening strategies. JAMA 1995;274:881–884.
39. Kerlikowske K: Presentation to the Board of Scientific Counselors meeting on October 22, 1993, Bethesda, MD.
40. Tabar L, Larson LG, Andersson I, et al.: Breast cancer screening with mammography in women aged 40–49. Report of the Organizing Committee and Collaborators, Falun Meeting, Falun, Sweden, March 21–22, 1996. Int J Cancer 1996;68:693–699.

41. Panel Statement—National Institutes of Health Consensus Development Conference Statement—Breast Cancer Screening for Women Ages 40–49. Bethesda, MD, January 23, 1997.
42. Bjurstam N, Bjorneld L, Duffy S, et al.: The Gothenberg Breast Cancer Screening Trial: Preliminary results on breast cancer mortality for women ages 39–49. Monog Natl Cancer Inst 1997;22: 87–92.
43. Andersson I, Janzon L: Reduced breast cancer mortality in women under 50: Update from the Malmo Mammographic Screening Program. Monog Natl Cancer Inst 1997;22:63–67.
44. Larson L-G, Andersson I, Bjurstam N, et al.: Updated overview of the Swedish randomized trials on breast cancer screening with mammography: Age group 40–49 at randomization. Monog Natl Cancer Inst 1997;22:57–61.
45. Hendrick RE, Smith RA, Rutledge JH, Smart CR: Benefit of screening mammography in women ages 40–49: A new meta-analysis of randomized controlled trials. Monog Natl Cancer Inst 1997;22:87–92.
46. Adami H, Malker B, Holmberg L, et al.: The relation between survival and age at diagnosis in breast cancer. N Engl J Med 1986;315:559–563.
47. Kerlikowske K, Grady D, Barclay J, et al.: Likelihood ratios for modern screening mammography: Risk of breast cancer based on age and mammographic interpretation. JAMA 1996;276:39–43.
48. Linver MN, Paster SB: Mammography outcomes in a practice setting by age: Prognostic factors, sensitivity, and positive biopsy rate. Monog Natl Cancer Inst 1997;22:113–117.
49. Shapiro S: Periodic screening for breast cancer: The HIP Randomized Controlled Trial. Monog Natl Cancer Inst 1997;22:27–30.
50. Tabar L, Duffy SW, Chen HH: Re: Quantitative interpretation of age-specific mortality reductions from the Swedish breast cancer screening trials. J Natl Cancer Inst 1996;88:52–53.
51. Bjurstam N, Bjorneld L, Duffy S, et al.: The Gothenberg Breast Cancer Screening Trial: Preliminary results on breast cancer mortality for women ages 39–49. Monog Natl Cancer Inst 1997;22: 87–92.
52. Sullivan DC, Zern RT: National Institutes of Health Consensus Development Conference Statement: Breast cancer screening for women ages 40–49. Minority report. Monogr Natl Cancer Inst 1997;22:xii.
53. Lachin JM: Introduction to sample size determination and power analysis for clinical trials. Controlled Clin Trials 1981;2:93–113.
54. Moher D, Dulberg C, Wells GA: Statistical power, sample size, and their reporting in randomized controlled trials. JAMA 1994; 272:122–124.
55. Prorok PC, Hankey BF, Bundy BN: Concepts and problems in the evaluation of screening programs. J Chron Dis 1981;34:159–171.
56. Fletcher SW: Whither scientific deliberation in health policy recommendations. N Engl J Med 1997;336:1180–1183.
57. Pauker SG, Kassirere JP: Contentions screening decisions: Does the choice matter? N Engl J Med 1997;336:1243–1244.
58. Moskowitz M: Breast cancer: Age-specific growth rates and screening strategies. Radiology 1986;161:37–41.
59. Tabar L, Faberberg G, Day NE, Holmberg L: What is the optimum interval between screening examinations:—An analysis based on the latest results of the Swedish two-county Breast Cancer Screening Trial. Br J Cancer 1987;55:547–551.
60. Kerlikowske K, Grady D, Barclay J, et al.: Effect of age, breast density, and family history on the sensitivity of first screening mammography. JAMA 1996;276:33–38.

APPENDIX
Unplanned Subgroup Analysis

The basis of the screening controversy was the use of unplanned subgroup analysis that separated the results for women aged 40–49 for separate analysis. Since none of the trials (including the Canadian trial) had sufficient numbers of women in these age groups to permit an anticipated 25% mortality reduction with statistical significance, these analyses were preordained to be misleading. Even when all the trials were grouped together, they had only one third the number of women needed for a benefit to be statistically significant. This would be similar to undertaking a poll to predict who will win the next Republican primary for President. A poll that included 1,000 individuals would likely be fairly predictive. However, if it turned out that only 100 of those polled were Republicans (only Republicans can vote in the Republican primary), it would be obvious that the sample size was too small to be predictive. This is equivalent to unplanned subgroup analysis that lacks statistical power. This is exactly what was permitted in the analysis of the early trial data.

Why Are Randomized, Controlled Trials Needed, and How Do They Work?

The best example of why RCTs are needed, and of how they work, can be seen in the following model. Assume that there are two women (A and B) and that they are identical in every way, such that the same cancer will develop in both women in the same breast at the same time and that the cancers will grow at the same rate. Five years after the first cancer cell was initiated, woman A has a mammogram and has her cancer detected. She is treated, but succumbs to breast cancer 10 years later. Woman B waits until she feels her cancer 7 years after it started. She is diagnosed and treated, but dies 8 years later from breast cancer. A review of survival information (date of diagnosis to date of death shows that woman A lived 10 years from the diagnosis of her breast cancer, while woman B only survived 8 years. Based on survival, it would appear that woman A benefited from early detection. This is misleading, however, in that they both died 15 years after the first cell became malignant. We knew about the cancer in the first woman earlier, but she still died at the same time as her twin. This is termed ‘‘leadtime bias.’’

The only way to have shown a benefit for woman A would have been if she had actually outlived woman B. By delaying her mortality, woman A would have shown the benefit from earlier detection, and lead time bias and other biases would have no influence on the result.

Since it is impossible to identify directly women who are cancer twins, large trials are needed with women divided randomly into two groups. If the number of participants is large enough, each woman in the screened group will, on a statistical basis, have her ‘‘twin’’ in the control group who is destined to die at the same time from the same type of cancer. If there is a benefit from screening, then fewer women will die each year, from breast cancer, in the screened group than the control group. For a benefit to appear for a woman in the screened group, however, her counterpart in the control group must die of breast cancer. This is the basis of

randomized, controlled trials. It is important to understand that if the control women do not die rapidly, the benefit for the screened women may not appear for many years.

The Benefit for Women in Their Forties Is Not Due to Their Having Reached the Age of 50 During the Trials and Screening Suddenly Beginning to Work

Understanding the basis of RCT is critical for understanding the results of RCT. For example, it has been suggested that the benefit that has been demonstrated for women aged 40–49 is due to their having reached the age of 50 and screening suddenly beginning to work. If there were some biological clock that told the body and cancers when the age of 50 (not menopause, but age 50) had been reached and that cancers could suddenly be detected and cured, then this explanation would be possible. There are several weaknesses and fallacies in the argument. The first is that it has no biological basis. Secondly, there is no proof that screening suddenly begins to work at the age of 50. Furthermore, analyzing trial data by the age at diagnosis, rather than the age at allocation, introduces significant biases. The misleading effects of analyzing data by the age at diagnosis can be seen by returning to the basic mechanism of pared twins in RCTs. Assume that twin A (in the screened group) has her cancer detected at the age of 48 and, consequently, does not die of breast cancer, while her twin (woman B) does not have her cancer detected until she is 51 and she dies at age 56. If the data are analyzed by the age at diagnosis (40–49 vs. 50 and older), woman A will be attributed to the younger group, but, since she has no counterpart in the control group, her failure to die from her cancer will be attributed as a nonlethal cancer under age 50 (and thus no benefit), while her twin's death will be attributed to the older group, and will increase the appearance of a benefit for women ages 50 and older. The only way to avoid this misleading analysis is to use the age at allocation, which is the way trial data are supposed to be analyzed and for which the benefit has been demonstrated.